Title	JC and Polytechnic Mathematics – Statistics
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic Liu Hui Ling, Ngee Ann Polytechnic
Date	17/6/2018

Page Number	Topic	JC H1	JC H2	Poly (Engineering and IT Courses)	Poly (Science Courses)
2	Discrete Probability Distribution	✓	✓	✓	✓
3	Binomial Distribution	✓	✓	✓	✓
8	Poisson Distribution		✓	✓	
11	Continuous Probability Distribution	✓	✓	✓	✓
13	Normal Distribution	✓	✓	✓	✓

Applicable to the following group of students.

- ✓ JC H1 – ‘A’ Level H1 Examination
- ✓ JC H2 – ‘A’ Level H2 Examination
- ✓ Polytechnic Engineering Courses – Engineering Mathematics – Statistical Analysis
- ✓ Polytechnic Information Technology Courses – Computing Mathematics – Statistical Analysis
- ✓ Polytechnic Science Courses – Biostatistics
- ✓ Polytechnic Business Courses – Business Statistics

Goals of this project

- ✓ Demonstrating the compatibility of Mathematics Syllabus of Junior College and Polytechnic.
- ✓ Promote the idea of sharing resources between students from both types of schools
- ✓ Promote the idea of collaborative revision between students from both type of schools.

This list doesn’t end here. There are many ways students from both types of schools can collaborate and it is up to the student to decide how they want it to be done.

Title	Concept of Discrete Probability Distribution and Discrete Random Variables
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic
Date	14/6/2018

Discrete Probability Distribution

- A category of probability distribution
- Takes in variables which are discrete.

Discrete Random Variables (Properties)

- Must be a non-negative integer (Example, the number of trials made. It would not make sense to say the number of trials made is a negative value, or a value that has decimal places.)
- A finite number of values (Examples, the number of trials made can't possibly be infinite. It must somehow stop at some point.)
- The sum of probability of every single variable, when sum up shall exactly be equal to 1 (If the sum of probability doesn't equate to 1, the variable is discrete but not random.)

Example of a Discrete Probability Distribution

Number of Trials as given. [Discrete Random Variable]	1	2	3	4
Probability	0.5	0.25	0.1	0.15

- ✓ Sum of all Probability Equals to 1, $P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1$
- ✓ The number of trials (Or the Discrete Variable) doesn't contain neither negative integers nor fractions.

Other Discrete Random Variables I thought of

- Number of phone calls received by customer service on a given day
- Number of employee approaching the technical helpdesk for assistance

In this case, you could be finding

- Probability of receiving k number of calls on a given day
- Probability of k number of employee approaching the helpdesk at a specified time-interval

Title	Polytechnic and A Level H2 Mathematics (Statistics) Binomial Distribution
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club]
Date	9/6/2018

Applicable to the following levels

- ✓ School of Information Technology Students (Computing Mathematics)
- ✓ School of Engineering (Engineering Mathematics – Statistical Analysis)
- ✓ School of Business Management (Statistics – Business Statistics)
- ✓ School of Chemical and Life Sciences – Biostatistics
- ✓ JC/MI Students – H2 Mathematics – Statistics

Due to my school's syllabus, it may or may not cover everything required for H2 Mathematics. JC/MI students should see referring to this guide as a last resort if you still don't know the basics.

To use the binomial distribution, the following requirements must be met.

- ➢ There will only be 2 possible outcomes (Success/Failure, Yes/No, etc.)
- ➢ Each trial is an independent event (that is, will not affect the subsequent trial or be affected by past trial)

You must also know the following information or able to derive the following details

- ➢ You know the probability of each trial
- ➢ You are given the total number of trials and the number of trials the probability is being calculated for, which will be shown in notation form in the next few pages.

Formula for Binomial Distribution Probability Given as Follows

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

[It may be written slightly differently in other textbooks. But they should mean the same thing.]

Notation	Meaning
$P(X = k)$	Probability of each trial, if the number of trials you are finding the probability for is exactly equal to k. (This is also called the discrete random variable notation.)
$\binom{n}{k}$	The number of combinations you can arrange the probability of the two outcomes. $$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$$ n refers to total number of trials you are performing k refers to the total number of trials you are finding the probability value for an outcome.
p^k	The probability of the outcome you are finding for after k number of independent trials. [E.g. The outcome can be Success or Yes.]
$(1-p)^k$	The probability of the alternate outcome you are finding for after k number of independent trials. [If the outcome you are finding the probability for is Success or Yes, then the alternate outcome you are finding the probability for are Failure or No respectively.]
$X \sim B(n, p)$	X means the discrete random variable of X. $\sim B$ means to be binomially distributed n means the total number of independent trials p means the probability of getting an outcome you are finding for.

Formula for Analysis of Data Binomially Distributed	
Finding the Mean (μ) [Also called the Expected Value]	$\mu = np$
Formula for Variance (σ^2)	$\sigma^2 = np(1-p)$
Formula for Standard Deviation (σ)	$\sigma = \sqrt{np(1-p)}$

Useful Tips:

When doing practices questions in school or just doing assignments, you may not have the answer key to check if your answer is correct. Downloading and installing probability distribution calculator apps saves you the trouble from having to ask the teacher and friends on the answer.

Binomial Distribution Questions and Example

[Section I]: Basic Calculation

Q1: Given the following binomial distribution and information.

$$X \sim B(5,0.3)$$

Evaluate the following

(a) $P(X = 2)$
(b) $P(X < 2)$
(c) $P(X < 3)$
(d) $P(X \geq 2)$

Q1(a)

$$P(X = 2) = \binom{5}{2} 0.3^2(1 - 0.3)^{5-2} = 10(0.09)(0.343) = 0.3087$$

Q1(b)

$$P(X < 2) = P(X = 0) + P(X = 1)$$

$$P(X = 0) = \binom{5}{0} 0.3^0(1 - 0.3)^{5-0} = 1(0.3)^0(0.7)^{5-0} = 0.16807$$

$$P(X = 1) = \binom{5}{1} 0.3^1(1 - 0.3)^{5-1} = 5(0.3)^1(0.7)^{5-1} = 0.36015$$

$$P(X < 2) = 0.16807 + 0.36015 = 0.52822$$

Q1(c)

$$P(X < 3) = 1 - [P(X = 3) + P(X = 4) + P(X = 5)]$$

[Values of all probabilities in binomial distribution must sum up to 1]

**Use the method that require the least number of calculation.

$$P(X = 3) = \binom{5}{3} 0.3^3(1 - 0.3)^{5-3} = 10(0.027)(0.7)^2 = 0.1323$$

$$P(X = 4) = \binom{5}{4} 0.3^4(1 - 0.3)^{5-4} = 5(0.0081)(0.7) = 0.02835$$

$$P(X = 5) = \binom{5}{5} 0.3^5(1 - 0.3)^{5-5} = 1(0.00243)(1) = 0.00243$$

$P(X = 3) + P(X = 4) + P(X = 5) = 0.16308$

$P(X < 3) = 1 - 0.16308 = 0.83692$
Q1(d) [From Answers Derived in Q1(b)]
$P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - (0.16807 + 0.3015)$
$= 1 - 0.46458$
$= 0.53542$

Section II (Application of Binomial Distribution)
Q2
A survey indicates that 60% of the school's student population is interested to participate in an event. You randomly selected 7 students who had participated in the survey.

(a) Is binomial distribution suitable for this question, please justify your answer.
(b) Find the probability that exactly 4 students are interested in the event.
(c) Find the probability that at most 3 students are interested in the event.
(d) Find the expected value, standard deviation and variance of the distribution.

(a) Yes. Every student's interest in the event can be regarded as independent.
There are only two possible outcomes, either a "YES" or a "NO".

(b)
$$P(X = 4) = \binom{7}{4}(0.6)^4(1 - 0.6)^{7-4} = 35(0.1296)(0.064) = 0.290304$$

(c)
$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

$$P(X = 0) = \binom{7}{0}(0.6)^0(1 - 0.6)^{7-0} = 0.00164$$

$$P(X = 1) = \binom{7}{1}(0.6)^1(1 - 0.6)^{7-1} = 0.01720$$

$$P(X = 2) = \binom{7}{2}(0.6)^2(1 - 0.6)^{7-2} = 0.00741$$

$$P(X = 3) = \binom{7}{3}(0.6)^3(1 - 0.6)^{7-3} = 0.19354$$

$P(X \leq 3) = 0.00164 + 0.01720 + 0.00741 + 0.19354 = 0.21979$

(d)

$\mu = np$

$\mu = 0.6(7) = 4.2$

$\sigma^2 = np(1 - p) = 4.2(1 - 0.6) = 1.62$

$\sigma = \sqrt{1.62} = 1.2728$

Title	Polytechnic and JC H2 Mathematics – Poisson Distribution
Author	Liu Hui Ling, Ngee Ann Polytechnic (Assisted by Chen Xin Yi)
Date	10/6/2018

Applicable to the following levels and types of education institution

- ✓ JC/MI – H2 Mathematics (Statistics)
- ✓ Engineering, Physics, Chemistry and Biology – Statistical Calculations
- ✓ Information Technology – Data Analytics

Apart from studying Business related modules, we also do some Business Statistics Module which drew my interest in this topic of probability distribution. I think we can just get straight to the point and explain what are the prerequisite and reason for using of this type of probability distribution.

Purpose of Poisson Distribution is to

- Calculate the probability of an event happening in the subsequent intervals when the mean rate of occurrences per unit of interval is given.

Information needed

- Mean occurrence rate
- Unit of Intervals (Unit of interval is the key word here. Without this unit of intervals, it is highly likely that the use of Poisson Distribution cannot be justified. Unit of intervals can come in terms of the time-interval, area-interval, volume-interval and etc.)

Requirements

- Multiple events cannot happen simultaneously
- All events must be <u>independent</u> (i.e. Unaffected by past events and will not affect subsequent events)

Formula used (Explanation given at the next page)

$$P(X = k) = e^{-\mu} \left(\frac{\mu^k}{k!}\right)$$

$P(X = k)$	Meaning: The probability that the number of event occurrence within the unit interval being exactly equal to k.
μ	Meaning: The mean occurrence per unit interval. (Sometimes people just call it, Expected Value or Expected Mean)
e	Refers to the Euler Constant. Approximately 2.71828, rounded off to 6 significant figures (Most modern scientific calculators should have this functionality, just locate the button e^x or e)

Question 1

Given that the number of days snow will fall as recorded by a particular research station in Antarctica follows a Poisson Distribution with a mean rate of 1.8 days per month.

(a) Find the probability that on a particular month, there is no incident of snowing recorded at the research station.

(b) Find the probability that on a particular month, there will exactly be more than 2 days where snow is recorded at the research station.

Q1(a)

$$P(X = 0) = \frac{e^{-1.8}(1.8^0)}{0!} = 0.16530$$

Q1(b)

$$P(X = 0) = \frac{e^{-1.8}(1.8^0)}{0!} = 0.16530$$

$$P(X = 1) = e^{-1.8}\left(\frac{1.8^1}{1!}\right) = 0.29754$$

$$P(X = 2) = e^{-1.8}\left(\frac{1.8^2}{2!}\right) = 0.26778$$

$$P(X > 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$
$$= 1 - (0.16530 + 0.29754 + 0.26778)$$
$$= 1 - (0.73062)$$
$$= 0.26938$$

Title	Concept of Continuous Random Variables and Continuous Probability Distribution
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club]
Date	15/6/2018

We discussed in previous pages the use of Discrete Probability Distribution, which takes in discrete values (i.e. non-negative integer values) as it's variables. Now we will be discussing about Continuous Probability Distribution.

The chart illustrates the differences between both types of distribution, each under its respective column.

Discrete Probability Distribution	**Continuous Probability Distribution**
Finite number of possible values	Infinite number of possible values
The random variable used in the topic are discrete. (Non-Negative Integer)	The random variables used in this topic are continuous. (Meaning, can contain decimal places, and occasionally can be negative) (Examples includes, amount of money in people's bank accounts, can be counted as negative should there be an overdraft.)
Objective: Find the probability of a specific variable	Objective: Find the probability that can satisfy a range of variables.
Typical Data Representation: Bar Graphs	Typical Data Representation: Line Graph
Sum of all Probabilities of Discrete Random Variables is equal to 1	Probabilities of all Range of the Continuous Random Variable, up till infinity, adds up to 1. (In calculus terms, the integral from every range up till infinity and negative infinity shall sum up to 1)

In calculus terms, for a continuous probability distribution, the probability you are finding is the area under curve or the integral of the distribution graph. Whereas in discrete probability distribution, you are finding the height of the bar within the graph.

As mentioned above, in a Continuous Probability Distribution, you are finding probability for the range of continuous variable that can satisfy the question. This means that statements like $P(X = k)$ where k is any number, is not valid and useless (which will be elaborated in subsequent paragraphs) in this type of distribution. Instead, you will only see statements like the following.

$P(X \geq 3)$
$P(X \leq 4)$
$P(X < 4)$
$P(X > 5)$
$P(X < 6)\ OR\ P(X > 8)$

Examples of Continuous Random Variables

- Height of Students in a School Population
- Amount of Time Spent Completing a Task Within a Competition

In this case, the probability you are finding for could be

- Probability of finding a student who is more than 1.8 meters in height
- Probability that a contestant could complete the task within the specified time interval (Example: Between 15 to 25 minute)

And I need to point out that in a Continuous Probability Distribution, the probability of a Continuous Random Variable being exactly equal to some value is literally 0.

This means the following statements are mathematically equivalent in the context of a continuous probability distribution.
$P(X \geq 5)$ is the same as $P(X > 5)$
$P(X \leq 4)$ is the same as $P(X < 4)$

In the next few pages, we will be discussing about Normal Distribution, a type of Continuous Probability Distribution.

Title	Polytechnic and A Level H2 Mathematics (Statistics) – Normal Distribution
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club]
Date	15/6/2018

Applicable to the following levels

- ✓ School of Information Technology Students (Computing Mathematics)
- ✓ School of Engineering (Engineering Mathematics – Statistical Analysis)
- ✓ School of Business Management Students (Statistics – Business Statistics)
- ✓ School of Chemical and Life Science – Biostatistics
- ✓ JC/MI Students – H2 Mathematics – Statistics

Items needed to start the topic

- ✓ Standard Normal Table

(Recommended, print a Standard Normal Table to refer to while doing your homework and assignments, while there are literally thousands of them on the internet, best is get from your school teacher and keep it. I also recommend you upload a copy to a cloud disks, just in case you lose the Standard Normal Table, you restore them quickly and reprint them.)

SEAB do have a copy of Standard Normal Table on their website. With enough searching you should be able to find it.

My school also issues its own version of the standard normal table
(I have seen Standard Normal Table issued by other schools before, they have different way of expressing the value of area under curve and different numerical accuracy requirements.)

Table of Notation	
$\boldsymbol{X \sim N(\mu, \sigma^2)}$	This is how a Normally Distributed Variable should be written. This literally means, The variable X is to be normally distributed, with a mean of μ, and a variance of σ^2. (Replace the symbols with values as specified in the questions you are going to answer)
$\boldsymbol{Z \sim N(0, 1)}$	Standard Normal Distribution. With mean as 0 and a variance of 1. Since $\sqrt{1} = 1$, the standard deviation of the distribution is also 1 in the case of a Standardized Normal Distribution. In this case, Z is the number of standard deviations away from the mean, also called the Z-score.

Table of Formula
Formula for Standardization
$$Z \sim N(0,1) = \frac{X - \mu}{\sigma}$$

Properties of a Normal Distribution Curve.

- Mean, Median and Mode are all on the same value
- Symmetrical at mean*, implying the left side of the Normal Distribution has a total area of 0.5 and the right side of the Normal Distribution has a total area of 0.5 as well.

(This is important to know as I am aware that some standard normal table out there are not as straightforward, I have seen other schools' standard normal table that shows value of area under curve from the mean to the Z-score, the most common types, however, shows area from the left of the distribution to the mean and shows the area from the left of the distribution all the way to the right of the distribution.)

Example Questions
Example 1:
(Taken from Oxford University Lecture Notes)

The marks of 500 candidates in an examination are normally distributed with a mean of 45 marks and a standard deviation of 20 marks.
If 20% of the candidates obtained a distinction by scoring x marks or more, estimate the value of x.

Written in Normal Distribution Notation
$X \sim N(45, 20^2)$

$P(20\%\ of\ the\ candidates\ socring \geq x\ marks)$
$= P(80\%\ of\ candidates\ scoring \leq x\ marks)$

Within the Standard Normal Table, I will look for the probability value closest to 0.800
(In this case, the standard normal table doesn't have a value exactly equal to 0.800.)

It turned out, the standard normal table probability value closest to 0.800 is 0.7995, under $z = 0.84$

Given
$z = \frac{x-\mu}{\sigma}$
Applying Standardization Formula

$0.84 = \frac{x-45}{20}$

$20(0.84) = x - 45$

$16.8 = x - 45$

$x = 16.8 + 45 = 61.8$

Example 2: (Taken from Online Sources)
The daily revenue of a small restaurant is approximately normally distributed with a mean of $530 and a standard deviation of $120. To be in profit, a restaurant must receive at least $350.

Find the probability that the restaurant will be in profit on any given day.

Given
$$z = \frac{x-\mu}{\sigma}$$

Applying Standardization Formula
$$Z = \frac{350-530}{120}$$

$$Z = -1.5$$

(Looking for $Z = -1.50$ in the Standard Normal Table, turns out the probability is 0.9932)

Thus, the probability of getting $\geq \$350$ is 0.9332.